A PROTOTYPE GENERIC EXPERT SYSTEM ARCHITECTURE FOR DATA REVIEW/VALIDATION

John Determan and Greg Becker
Bechtel BWXT Idaho, LLC
Idaho National Engineering and Environmental Laboratory
P.O. Box 1625
Idaho Falls, ID 83415
jcd@inel.gov, gkb1@inel.gov

ABSTRACT

Nondestructive assay (NDA) waste characterization data generated for use in the National Transuranic Program must be of known and demonstrable quality. Each measurement is required to receive an independent technical review by a qualified expert. With hundreds of measures acquired on each drum, and hundreds of thousands of such drums in the national inventory, a large and costly data review/validation effort is mandated. For this reason, researchers at the Idaho National Engineering and Environmental Laboratory have developed a generic expert system framework for review/validation of nondestructive waste assay data, known as NDA-DRXS. This system expresses the process of data review in a set of expert system rules, and is designed around an object-oriented framework that can be easily customized for application at sites with different NDA configurations and waste inventory properties.

INTRODUCTION

Management of U.S. Department of Energy (DOE) defense generated containerized transuranic (TRU) waste requires determination of the entrained TRU mass and associated parameters. Nondestructive assay (NDA) techniques are the most common and efficient means to determine the TRU material quantity. quality assurance objectives (QAOs) for NDA techniques used to characterize TRU waste destined for the Waste Isolation Pilot Plant (WIPP) are delineated in the National TRU Program Transuranic Waste Characterization WIPP Waste Acceptance Criteria.[1]

Technically justifying compliance with applicable requirements and QAOs for TRU waste forms in the DOE inventory can be a complex process. Some waste form configurations manifest NDA system response complexities that diminish the ability to clearly establish compliance. Such complexities lead to the requirement that technical reviews be performed at the data generation level for each assay to ensure operational boundaries are maintained relative to QAPP requirements.

Technical review of waste NDA measurement data, though warranted with respect to present day waste NDA system capabilities, is labor intensive. Hence, it is desirable to have an automated system to perform the technical review. The automated system must be capable of providing a comprehensive waste assay data assessment, and must be reproducible, audible, and compatible with the overall throughput requirements of the waste characterization process. Therefore, an evaluation of expert system technology was undertaken, and a prototype data review expert system was designed and built.

Generally, similar measurement devices are employed at different sites where nondestructive waste assay (NDA) is performed. A common NDA system configuration is one that employs active and passive neutron counting techniques to estimate the mass of transuranic nuclides in a waste container. This configuration also requires radionuclidic and isotopic composition data. The most common means of acquiring this data is through high-resolution gamma spectroscopy. Another common NDA system configuration is one that employs passive high-resolution gamma spectroscopy combined with sample gamma transmission measurements. Extensive data is collected from these devices, and so a lengthy data reduction process is necessitated. Our system ensures that this data reduction process was performed correctly, and that the reduced data is consistent, believable, and representative of a valid set of measurements. The system may also be configured to examine raw data, however, the review processes we have modeled so far emphasize review of the reduced data.

ANALYSIS OF THE DATA REVIEW PROCESS

NDA experts at two different sites where the first of the two configurations mentioned above is in use were interviewed to discover the nature of their data review activities. The observed similarities and differences in their data review processes guided the development of a generic expert system framework for NDA data review.

At a high level, the similarities in the data review strategies were strong. Both organizations had similar goals that the technical data review was to achieve. Problems in the raw and reduced data could be categorized by level of severity, and each category entailed a specific requisite action, as shown in Table 1. Some types of problem required a specific response, such as repeating a measurement or calculation, or acknowledging that an assay would never be meaningful on a particular container. Other problems, however, could only be evaluated in the full context of the assay. Consideration of the data set as a whole allows flexibility in evaluating individual measurement parameter deficiencies. In such cases, questionable items are flagged, and the assay is set aside for later (human expert) analysis. Another strong

Table 1. Data review classification scheme.

Severity Level	Nature of Problem	Requisite Action
1	Value outside of technical envelope	Reject assay
2	Improper control setting, or other value	Repeat measurement
3	Improper constant or table used in data reduction	Repeat calculation
4	Data quality compromised	Flag data as suspicious, but proceed
5	None	Accept assay

similarity between the review processes at the two sites was the nature of the review process itself. Data validation rules are performed on a specific set of quantities, and the outcome of these validation procedures feeds into the final disposition of the assay.

At the lowest level, however, differences were also strong. Slight differences in hardware configuration, measurement technique, and data reduction methods, as well as differences in waste inventory characteristics, resulted in many validation procedures that were unique to a given waste generator site. Thus, a picture emerged of the need to develop a classification scheme, and simultaneously the need to implement site-specific validation rules, that fed their results into this scheme. The framework would be required to allow easy definition and integration of site-specific checks into the data validity classification scheme. Finally, it is important to stress that the framework consists of only the code needed to implement the common logic of the technical data review process. No site-specific validation rules are built into the framework itself, nor are they ever shared between different implementations of the expert system.

DESIGN OF A GENERIC EXPERT SYSTEM ARCHITECTURE FOR DATA REVIEW

The data produced by complex measurement systems may be viewed as a complete specification of the state of the measurement configuration. When this specification consists of hundreds of measured values, however, considerable effort is required to transform this information into a coherent picture (i.e., the measurements are self-consistent, complete, accurate, and believable, or some known and specific irregularity exists). An expert system can perform this analysis quickly and accurately. Measurement data to be reviewed by the expert system are represented as a set of facts. Facts are data structures that associate a symbolic name and a value (or set of values, in some cases). Expert systems are rule-based systems. This means that the operational logic of the system is expressed in terms of a series of rules of the form "IF *pattern* THEN *action*," where *pattern* represents some combination of conditions in the input facts. Typically, actions taken by the rules transform the facts represented in the system to a simpler form. This process is controlled by a component of the expert system known as an inference engine. The inference engine repeatedly applies the expert system rules to the known facts until the facts have been simplified to the point that no more actions can be taken. This reduced set of facts is still indicative of the state of the measurement configuration. These facts can then be used to produce a set of descriptive statements that inform human users of the expert system of the state of the measurement configuration.

An important step in producing a practical expert system is to model the expert technical review process in generic, systematic terms. Our initial motivation for producing a generic model was to ease the process of adapting a data review system to other types of measurement systems. But another important benefit of having a generic model is that the expert system is no longer designed around a specific review process. Review processes change, and redesigning the expert system to accommodate every process modification would not be feasible. In other words, a generic expert system is not only adaptable to new applications, but is much more robust in the face of an evolving review process.

Object-oriented programming languages offer powerful techniques for designing generic programs. The concepts of class inheritance and polymorphism are principal among these. Class inheritance allows attributes and functionality of physical objects to be modeled at different levels. Generic features of physical objects may be modeled in a so-called base class; specific types of these objects are modeled as derived classes. The derived classes inherit the attributes and functionality of the base class, but add their own specific attributes and functionality. On the one hand, there is a definite economy to modeling the common features of a set of similar objects only once, but the advantages of this approach go much further. Polymorphism is a feature of object-oriented languages that allows code to be written in terms of the base class, but applied without modification to any specific derived class object. Another way of saying this is that the base and derived classes share (at least in part) a common interface such that when functions are invoked on code written generically in terms of base class objects, it may be executed on the more specific derived class objects. The end result is that a generic process may be coded once and easily extended to new situations by deriving specific classes to model the particular elements of a new application.

The expert system language Clips is a rule-based language that incorporates object-oriented features.[2] Thus, in Clips, objects may be used in place of simple facts to represent input to the expert system. The rules of the expert system express the common logic of data review, and are written in terms of base class objects that model the common features of measurement data. Specific implementations of the expert system are produced by developing a set of derived classes to model the particular measurements to be reviewed in a given measurement configuration. The system was implemented within a C++ user-interface running under Windows NT.

Figure 1 diagrams a framework that accomplishes the goals specified above. The object model defines components for modeling site-specific data validation rules, messages, and lists used to control the data review process. The rule module captures the common logic of the data review process. The module is divided into specific phases of operation that include system initialization, item verification, message generation, and message output. The rule and object modules are discussed in detail in the following sections.

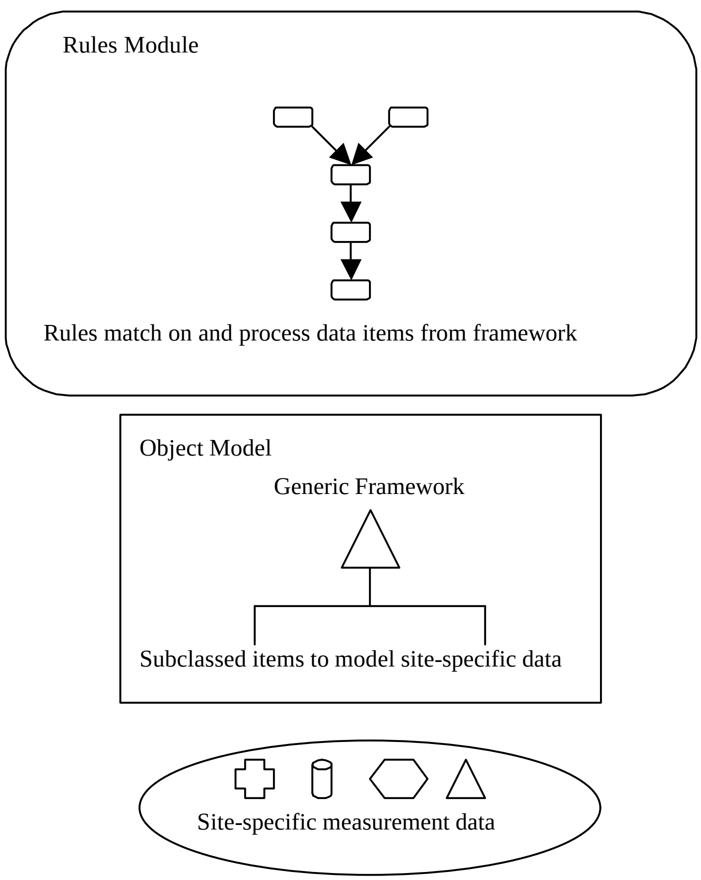

Figure 1. Overview of NDA-DRXS architecture.

Object Model

There are four base classes in the object model, from which all other classes are derived (excluding the interface classes such as file parsers, dialog windows, display windows, etc.), as shown in Figure 2. The four base classes are (1) List, (2) Summary-list, (3) Item-list, and (4) Data-item. These are the base classes used to model any piece of data that needs to be validated. The measured quantities represented by the Data-item class are divided into subgroups where each subgroup has a common validation rule. These validation rules may be as simple as checking that a Data-item's value is not equal to zero, or as involved as confirming statistical consistency with one or more other Data-items. It sometimes occurs that not all of the data required by a validation rule is available, or preconditions on components of a validation rule may not be met. Further processing is prevented in this case, and a message is generated indicating that a particular test could not be performed. Depending on the result of the validation rule, a message indicating success or failure of the test will be generated.

The List base class provides all of the usual features and operations of the single-linked list data structure (e.g., add items to the list, remove items from the list). Item-list and Summary-list are derived from List, and each serves a purpose specific to the data review system. Specific lists are derived from Item-list to represent groups of Data-items with the same validation rule. Every class derived from Data-item has a corresponding class derived from Item-list. As each Data-item is verified it is removed from the corresponding Item-list, such that an empty Item-list indicates that no problems were encountered.

Several classes are derived from Summary-list that are part of the generic framework. Each particular Summary-list holds the names of Item-lists for which the failure of the related validation rule results in a common problem classification, and can form a message pertaining to the problem classification (i.e., "Assay is rejected"). Just as an empty Item-list indicates that no failures occurred with respect to a given validation rule, an empty Summary-list indicates that no validation rules resulting in a particular problem were encountered.

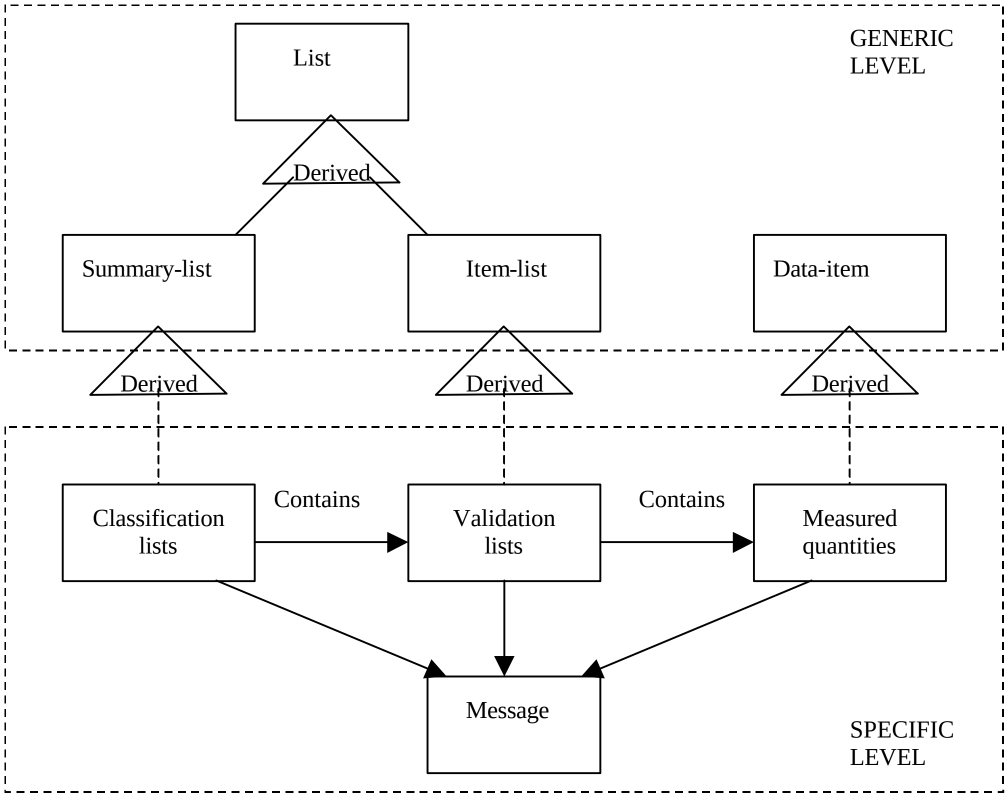

Figure 2. Object relationship diagram of the generic expert system for NDA data review.

Rule Module

Assays are typically performed in daily batches. Batches include replicate assays and assays performed on drums with known contents, according to quality assurance procedures. Validation rules can be performed on a whole batch of assays, or on a single assay. As shown in Figure 3, the rule module is organized as a series of phases of operation. While the figure is somewhat abbreviated, it shows the principal phases of the rule module: batch- and assay-initialization, verification, generate-output, and output. Rules for initiating data review detect whether batch characteristics or an individual assay are being validated and initiate operation within the appropriate phase for batch or assay initialization. The initialization phases populate the Item- and Summary-lists used to control the review process.

When all system objects are initialized, processing enters the verification phase. During verification, Data-item objects first confirm that sufficient data is available to perform the associated validation rule. Failure to confirm that sufficient data is available for validation results in a message to the user to that effect. Otherwise, Data-item objects perform their validation rules, and those that succeed form

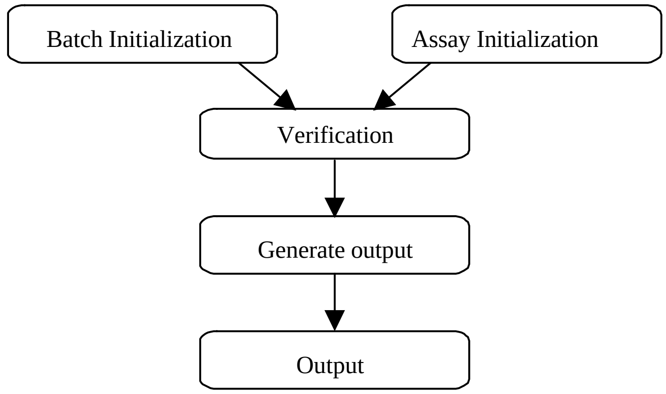

Figure 3. Schematic of NDA data review process.

appropriate messages. By the conclusion of this phase, all Data-item objects have formed a message concerning either the success or inapplicability of their validation rule. Data-item objects that have not been validated remain in their respective Item-lists. Empty Item-lists are destroyed, and their names removed from Summary-lists.

In the phase Generate Output, non-empty Summary-lists indicate a particular problem has occurred, and an appropriate message is generated. Likewise, the remaining Item-lists roll through their contents and generate failure messages for each named item.

In the Output phase, all message objects are directed to the appropriate output destination(s), and printed in the appropriate order. There are a very large number of checks being performed, such that the success messages would overwhelm the reviewer. Consequently, only a short report is directed to a window in the user interface. The short report omits the success messages. A long report, omitting nothing, is directed to a text file. The short report contains sufficient information for a knowledgeable user to interpret the results. The long report is useful as an autidable paper trail, and is also useful during system verification. The long and short reports follow the format shown in Figure 4.

IMPLEMENTING A REVIEW SYSTEM

The architecture described above allows for the easy specification of a review system. This is due to the fact that rule development has been localized within a single object, and such objects are integrated into a

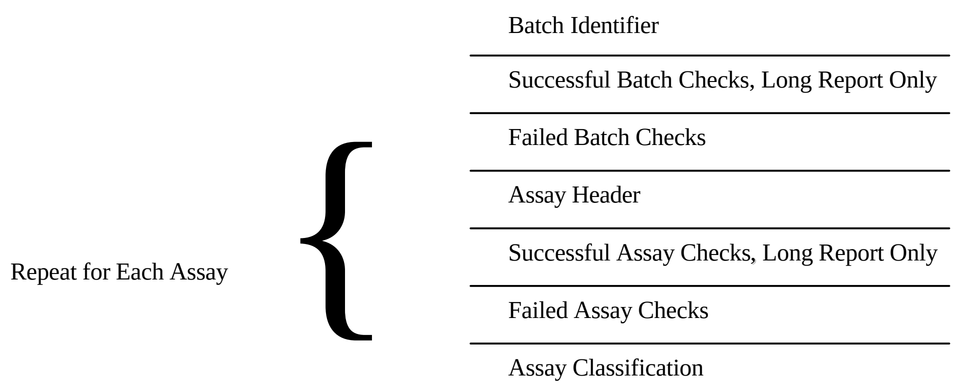

Figure 4. Output ordering.

functioning expert system automatically by the combined rule/object framework. The developer need only focus on crafting objects within the framework, instead of developing chains of rules that are both self-consistent and non-interfering with other rule chains. The order imposed by the framework also speeds the debugging process considerably.

Another, only partially realized possibility opened by the framework, is the possibility to create an editor/compiler tool that would allow a non-specialist to develop a review system. At present a dialog-based application exists for collecting data from the developer and producing the set of inputs that represent a particular rule, but it is still required that some of this input is directly entered as Clips code. It is reasonable to expect that requirements could be entered in some more natural form, and in turn be compiled to Clips code.

Definition of a set of Summary-lists is the first step in implementing a review system. For non-destructive waste assay, a set of generally applicable Summary-lists is already defined in Table 1. These objects represent user-defined classifications that are a principal outcome of the review process. In a particular implementation, some of these may not be needed, and others may need to be defined. But once the need for a particular classification has been defined, implementation represents only a few minutes effort.

Elicitation of validation rules is the next step. Rules are specified by relating them on the one hand to the classification their failure entails, and on the other to particular Data-items to be checked by the validation rule. An Item-list is derived that names the Summary-list with which the rule is associated, and contains the names of Data-items to be checked by the rule. A Data-item is derived with specific methods that implement the validation rule in question. More complicated validation rules may involve rule chaining, that is, their input is the output of other object's validation rules.

SUMMARY AND CONCLUSIONS

An expert system for nondestructive assay data review has been developed that is easily adaptable to different measurement system configurations.

- Easy adaptability to site-specific data requirements has been achieved through the combination of rule-based and object-oriented programming techniques.
- The common logic of data review has been captured in a simple set of expert system rules.
- An object-oriented framework models generic aspects of the data to be verified, and is adapted to a specific waste generation site through the object-oriented techniques of inheritance and polymorphism.

ACKNOWLEDGMENTS

This document was prepared for the U.S. Department of Energy Assistant Secretary for Environmental Management under DOE Idaho Operations Office Contract DE-AC07-99ID13727.

REFERENCES

1. U.S. Department of Energy Carlsbad Area Office, *Contact Handled Transuranic Waste Acceptance Criteria for the Waste Isolation Pilot Plant*, Revision 7, Technical Report DOE/WIPP-069, U.S. DOE Carlsbad Area Office, 1999.

2. J. C. Giarratano, *Clips User's Guide*, Technical Report, JSC-25013, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, 1999.